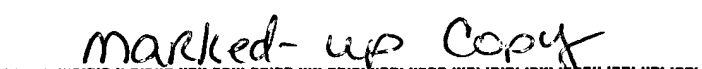
marked-up Copy



US 20030113887A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2003/0113887 A1**

Dujon et al. (43) **Pub. Date: Jun. 19, 2003**

(54) **NUCLEOTIDE SEQUENCE ENCODING THE ENZYME I-SCEI AND THE USES THEREOF**

(75) Inventors: **Bernard Dujon**, Gif Sur Yvette (FR); **Andre Choulika**, Paris (FR); **Laurence Colleaux**, Edinburgh (GB); **Cecile Fairhead**, Malakoff (FR); **Arnaud Perrin**, Paris (FR); **Anne Plessis**, Paris (FR); **Agnes Thierry**, Paris (FR)

Correspondence Address:
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER LLP
1300 I STREET, NW
WASHINGTON, DC 20005 (US)

(73) Assignee: **Institut Pasteur and Universite Paris VI**, both of Paris (FR)

(21) Appl. No.: **09/836,169**

(22) Filed: **Apr. 18, 2001**

Related U.S. Application Data

(60) Continuation of application No. 09/196,131, filed on Nov. 20, 1998, now Pat. No. 6,238,924, which is a continuation of application No. 08/417,226, filed on Apr. 5, 1995, now Pat. No. 5,962,327, which is a division of application No. 07/971,160, filed on Nov. 5, 1992, now Pat. No. 5,474,896, which is a continuation-in-part of application No. 07/879,689, filed on May 5, 1992, now abandoned.

Publication Classification

(51) **Int. Cl.**[7] **C12N 9/00**; C07H 21/04; C12P 21/02; C12N 5/06

(52) **U.S. Cl.** **435/183**; 435/325; 435/320.1; 435/69.1; 536/23.2

(57) **ABSTRACT**

An isolated DNA encoding the enzyme I-SceI is provided. The DNA sequence can be incorporated in cloning and expression vectors, transformed cell lines and transgenic animals. The vectors are useful in gene mapping and site-directed insertion of genes.

2670

```
                                        ATG CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG  2672
                                        M   H   M   K   N   I   K   K   N   Q   V   M    12

2671 AAC CTC GGT CCG AAC TCT AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC  2730
  13 N   L   G   P   N   S   K   L   L   K   E   Y   K   S   Q   L   I   E   L   N    32

2731 ATC GAA CAG TTC GAA GCA GGT ATC GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT  2790
  33 I   E   Q   F   E   A   G   I   G   L   I   L   G   D   A   Y   I   R   S   R    52

2791 GAT GAA GGT AAA ACC TAC TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC  2850
  53 D   E   G   K   T   Y   C   M   Q   F   E   W   K   N   K   A   Y   M   D   H    72

2851 GTA TGT CTG CTG TAC GAT CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC  2910
  73 V   C   L   L   Y   D   Q   W   V   L   S   P   P   H   K   K   E   R   V   N    92

2911 CAC CTG GGT AAC CTG GTA ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC  2970
  93 H   L   G   N   L   V   I   T   W   G   A   Q   T   F   K   H   Q   A   F   N   112

2971 AAA CTG GCT AAC CTG TTC ATC GTT AAC AAC AAA AAA ACC ATC CCG AAC AAC CTG GTT GAA  3030
 113 K   L   A   N   L   F   I   V   N   N   K   K   T   I   P   N   N   L   V   E   132

3031 AAC TAC CTG ACC CCG ATG TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT  3090
 133 N   Y   L   T   P   M   S   L   A   T   W   F   M   D   D   G   G   K   W   D   152

3091 TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA  3150
 153 Y   N   K   N   S   T   N   K   S   I   V   L   N   T   Q   S   F   T   F   E   172

3151 GAA GTA GAA TAC CTG GTT AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC GTA AAA  3210
 173 E   V   E   Y   L   V   K   G   L   R   N   K   F   Q   L   N   C   Y   V   K   192

3211 ATC AAC AAA AAC AAA CCG ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC  3270
 193 I   N   K   N   K   P   I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N   212

3271 CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC  3330
 213 L   I   K   P   Y   L   I   P   Q   M   M   Y   K   L   P   N   T   I   S   S   232

3331 GAA ACT TTC CTG AAA TAA
 233 E   T   F   L   K   *
```

[0013] This invention also relates to a DNA sequence comprising a promoter operatively linked to the DNA sequence of the invention encoding the enzyme I-SceI.

[0014] This invention further relates to an isolated RNA complementary to the DNA sequence of the invention encoding the enzyme I-SceI and to the other DNA sequences described herein.

[0015] In another embodiment of the invention, a vector is provided. The vector comprises a plasmid, bacteriophage, or cosmid vector containing the DNA sequence of the invention encoding the enzyme I-SceI.

[0016] In addition, this invention relates to *E. coli* or eukaryotic cells transformed with a vector of the invention.

[0017] Also, this invention relates to transgenic animals containing the DNA sequence encoding the enzyme I-SceI and cell lines cultured from cells of the transgenic animals.

[0018] In addition, this invention relates to a transgenic organism in which at least one restriction site for the enzyme I-SceI has been inserted in a chromosome of the organism.

[0019] Further, this invention relates to a method of genetically mapping a eukaryotic genome using the enzyme I-SceI.

[0020] This invention also relates to a method for in vivo site directed recombination in an organism using the enzyme I-SceI.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] This invention will be more fully described with reference to the drawings in which:

[0022] **FIG. 1** depicts the universal code equivalent of the mitochondrial I-SceI gene.

[0023] **FIG. 2** depicts the nucleotide sequence of the invention encoding the enzyme I-SceI and the amino acid sequence of the natural I-SceI enzyme.

[0024] **FIG. 3** depicts the I-SceI recognition sequence and indicates possible base mutations in the recognition site and the effect of such mutations on stringency of recognition.

[0025] **FIG. 4** is the nucleotide sequence and deduced amino acid sequence of a region of plasmid pSCM525. The nucleotide sequence of the invention encoding the enzyme I-SceI is enclosed in the box.

[0026] **FIG. 5** depicts variations around the amino acid sequence of the enzyme I-SceI.

[0027] **FIG. 6** shows Group I intron encoding endonucleases and related endonucleases.

[0028] **FIG. 7** depicts yeast expression vectors containing the synthetic gene for I-SceI.

[0029] **FIG. 8** depicts the mammalian expression vector PRSV I-SceI.

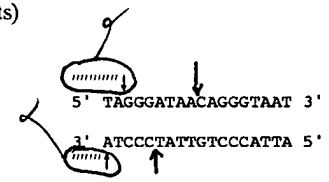
[0030] **FIG. 9** is a restriction map of the plasmid pAF100. (See also YEAST, 6:521-534, 1990, which is relied upon and incorporated by reference herein).

[0048] Cations: Enzymatic activity requires Mg^{++} (8 mM is optimum). Mn^{++} can replace Mg^{++}, but this reduces the stringency of recognition.

[0049] Optimum conditions for activity: high pH (9 to 10), temperature 20-40° C., no monovalent cations.

[0050] Enzyme stability: I-SceI is unstable at room temperature. The enzyme-substrate complex is more stable than the enzyme alone (presence of recognition sites stabilizes the enzyme.)

[0051] The enzyme I-SceI has a known recognition site. (ref. 14.) The recognition site of I-SceI is a non-symmetrical sequence that extends over 18 bp as determined by systematic mutational analysis. The sequence reads: (arrows indicate cuts)



[0052] The recognition site corresponds, in part, to the upstream exon and, in part, to the downstream exon of the intron plus form of the gene.

[0053] The recognition site is partially degenerate: single base substitutions within the 18 bp long sequence result in either complete insensitivity or reduced sensitivity to the enzyme, depending upon position and nature of the substitution.

[0054] The stringency of recognition has been measured on:

[0055] 1—mutants of the site.

[0056] 2—the total yeast genome (*Saccharomyces cerevisiae*, genome complexity is 1.4×10^7 bp). Data are unpublished.

[0057] Results are:

[0058] 1—Mutants of the site: As shown in **FIG. 3**, there is a general shifting of stringency, i.e., mutants severely affected in Mg^{++} become partially affected in Mn^{++}, mutants partially affected in Mg^{++} become unaffected in Mn^{++}.

[0059] 2—Yeast: In magnesium conditions, no cleavage is observed in normal yeast. In the same condition, DNA from transgenic yeasts is cleaved to completion at the artificially inserted I-SceI site and no other cleavage site can be detected. If magnesium is replaced by manganese, five additional cleavage sites are revealed in the entire yeast genome, none of which is cleaved to completion. Therefore, in manganese the enzyme reveals an average of 1 site for ca. 3 millions based pairs ($5/1.4\times10^7$ bp).

[0060] Definition of the recognition site: important bases are indicated in **FIG. 3**. They correspond to bases for which severely affected mutants exist. Notice however that:

[0061] 1—All possible mutations at each position have not been determined; therefore a base that does not correspond to a severely affected mutant may still be important if another mutant was examined at this very same position.

[0062] 2—There is no clear-cut limit between a very important base (all mutants are severely affected) and a moderately important base (some of the mutants are severely affected). There is a continuum between excellent substrates and poor substrates for the enzyme.

[0063] The expected frequency of natural I-SceI sites in a random DNA sequence is, therefore, equal to $(0.25)^{-18}$ or (1.5×10^{-11}) In other words, one should expect one natural site for the equivalent of ca. 20 human genomes, but the frequency of degenerate sites is more difficult to predict.

[0064] I-SceI belongs to a "degenerate" subfamily of the two-dodecapeptide family. Conserved amino acids of the dodecapeptide motifs are required for activity. In particular, the aspartic residues at positions 9 of the two dodecapeptides cannot be replaced, even with glutamic residues. It is likely that the dodecapeptides form the catalytic site or part of it.

[0065] Consistent with the recognition site being non-symmetrical, it is likely that the endonucleolytic activity of I-SceI requires two successive recognition steps: binding of the enzyme to the downstream half of the site (corresponding to the downstream exon) followed by binding of the enzyme to the upstream half of the site (corresponding to the upstream exon). The first binding is strong, the second is weaker, but the two are necessary for cleavage of DNA. In vitro, the enzyme can bind the downstream exon alone as well as the intron-exon junction sequence, but no cleavage results.

[0066] The evolutionarily conserved dodecapeptide motifs of intron-encoded I-SceI are essential for endonuclease activity. It has been proposed that the role of these motifs is to properly position the acidic amino acids with respect to the DNA sequence recognition domains of the enzyme for the catalysis of phosphodiester bond hydrolysis (ref. P3).

[0067] The nucleotide sequence of the invention, which encodes the natural I-SceI enzyme is shown in **FIG. 2**. The nucleotide sequence of the gene of the invention was derived by dideoxynucleotide sequencing. The base sequences of the nucleotides are written in the 5'→3' direction. Each of the letters shown is a conventional designation for the following nucleotides:

A	Adenine
G	Guanine
T	Thymine
C	Cytosine.

[0068] It is preferred that the DNA sequence encoding the enzyme I-SceI be in a purified form. For instance, the sequence can be free of human blood-derived proteins, human serum proteins, viral proteins, nucleotide sequences encoding these proteins, human tissue, human tissue components, or combinations of these substances. In addition, it is preferred that the DNA sequence of the invention is free of extraneous proteins and lipids, and adventitious micro-

organisms, such as bacteria and viruses. The essentially purified and isolated DNA sequence encoding I-SceI is especially useful for preparing expression vectors.

[0069] Plasmid pSCM525 is a pUC12 derivative, containing an artificial sequence encoding the DNA sequence of the invention. The nucleotide sequence and deduced amino acid sequence of a region of plasmid pSCM525 is shown in **FIG. 4**. The nucleotide sequence of the invention encoding I-SceI is enclosed in the box. The artificial gene is a BamHI—SalI piece of DNA sequence of 723 base pairs, chemically synthesized and assembled. It is placed under tac promoter control. The DNA sequence of the artificial gene differs from the natural coding sequence or its universal code equivalent described in Cell (1986), Vol. 44, pages 521-533. However, the translation product of the artificial gene is identical in sequence to the genuine omega-endonuclease except for the addition of a Met-His at the N-terminus. It will be understood that this modified endonuclease is within the scope of this invention.

[0070] Plasmid pSCM525 can be used to transform any suitable *E. coli* strain and transformed cells become ampicillin-resistant. Synthesis of the omega-endonuclease is obtained by addition of I.P.T.G. or an equivalent inducer of the lactose operon system.

[0071] A plasmid identified as pSCM525 containing the enzyme I-SceI was deposited in *E. coli* strain TG1 with the Collection Nationale de Cultures de Microorganismes (C.N.C.M.) of Institut Pasteur in Paris, France on Nov. 22, 1990, under culture collection deposit Accession No. I-1014. The nucleotide sequence of the invention is thus available from this deposit.

[0072] The gene of the invention can also be prepared by the formation of 3'→5' phosphate linkages between nucleoside units using conventional chemical synthesis techniques. For example, the well-known phosphodiester, phosphotriester, and phosphite triester techniques, as well as known modifications of these approaches, can be employed. Deoxyribonucleotides can be prepared with automatic synthesis machines, such as those based on the phosphoramidite approach. Oligo- and polyribonucleotides can also be obtained with the aid of RNA ligase using conventional techniques.

[0073] This invention of course includes variants of the DNA sequence of the invention exhibiting substantially the same properties as the sequence of the invention. By this it is meant that DNA sequences need not be identical to the sequence disclosed herein. Variations can be attributable to single or multiple base substitutions, deletions, or insertions or local mutations involving one or more nucleotides not substantially detracting from the properties of the DNA sequence as encoding an enzyme having the cleavage properties of the enzyme I-SceI.

[0074] **FIG. 5** depicts some of the variations that can be made around the I-SceI amino acid sequence. It has been demonstrated that the following positions can be changed without affecting enzyme activity:

positions −1 and −2 are not natural. The two amino acids are added due to cloning strategies.	
positions 1 to 10:	can be deleted.
position 36:	G is tolerated.
position 40:	M or V are tolerated.
position 41:	S or N are tolerated.
position 43:	A is tolerated.
position 46:	V or N are tolerated.
position 91:	A is tolerated.
positions 123 and 156:	L is tolerated.
position 223:	A and S are tolerated.

[0075] It will be understood that enzymes containing these modifications are within the scope of this invention.

[0076] Changes to the amino acid sequence in **FIG. 5** that have been demonstrated to affect enzyme activity are as follows:

position 19:	L to S
position 38:	I to S or N
position 39:	G to D or R
position 40:	L to Q
position 42:	L to R
position 44:	D to E, G or H
position 45:	A to E or D
position 46:	Y to D
position 47:	I to R or N
position 80:	L to S
position 144:	D to E
position 145:	D to E
position 146:	G to E
position 147:	G to S

[0077] It will also be understood that the present invention is intended to encompass fragments of the DNA sequence of the invention in purified form, where the fragments are capable of encoding enzymatically active I-SceI.

[0078] The DNA sequence of the invention coding for the enzyme I-SceI can be amplified in the well known polymerase chain reaction (PCR), which is useful for amplifying all or specific regions of the gene. See e.g., S. Kwok et al., J. Virol., 61:1690-1694 (1987); U.S. Pat. Nos. 4,683,202; and 4,683,195. More particularly, DNA primer pairs of known sequence positioned 10-300 base pairs apart that are complementary to the plus and minus strands of the DNA to be amplified can be prepared by well known techniques for the synthesis of oligonucleotides. One end of each primer can be extended and modified to create restriction endonuclease sites when the primer is annealed to the DNA. The PCR reaction mixture can contain the DNA, the DNA primer pairs, four deoxyribonucleoside triphosphates, $MgCl_2$, DNA polymerase, and conventional buffers. The DNA can be amplified for a number of cycles. It is generally possible to increase the sensitivity of detection by using a multiplicity of cycles, each cycle consisting of a short period of denaturation of the DNA at an elevated temperature, cooling of the reaction mixture, and polymerization with the DNA polymerase. Amplified sequences can be detected by the use of a technique termed oligomer restriction (OR). See, R. K. Saiki et al., Bio/Technology 3:1008-1012 (1985).

[0079] The enzyme I-SceI is one of a number of endonucleases with similar properties. Following is a listing of related enzymes and their sources.

[0080] Group I intron encoded endonucleases and related enzymes are listed below with references. Recognition sites are shown in FIG. 6.

Enzyme	Encoded by	Ref
I-SceI	Sc LSU-1 intron	this work
I-SceII	Sc cox1-4 intron	Sargueil et al., NAR (1990) 18, 5659–5665
I-SceIII	Sc cox1-3 intron	Sargueil et al., MGG (1991) 225, 340–341
I-SceIV	Sc cox1-5a intron	Seraphin et al. (1992) in press
I-CeuI	Ce LSU-5 intron	Marshall, Lemieux Gene (1991) 104, 241–245
I-CreI	Cr LSU-1 intron	Rochaix (unpublished)
I-PpoI	Pp LSU-3 intron	Muscarella et al., MCB (1990) 10, 3386–3396
I-TevI	T4 td-1 intron	Chu et al., PNAS (1990) 87, 3574–3578 and Bell-Pedersen et al. NAR (1990) 18, 3763–3770.
I-TevII	T4 sunY intron	Bell-Pedersen et al. NAR (1990) 18, 3763–3770.
I-TevIII	RB3 nrdB-1 intron	Eddy, Gold, Genes Dev. (1991) 5, 1032–1041
HO	HO yeast gene	Nickoloff et al., MCB (1990) 10, 1174–1179
Endo SceI	RF3 yeast mito. gene	Kawasaki et al., JBC (1991) 266, 5342–5347

[0081] Putative new enzymes (genetic evidence but no activity as yet) are I-CsmI from cytochrome b intron 1 of *Chlamydomonas smithii* mitochondria (ref. 15), I-PanI from cytochrome b intron 3 of *Podospora anserina* mitochondria (Jill Salvo), and probably enzymes encoded by introns Nc nd1.1 and Nc cob.! from *Neurospora crassa*.

nd1.1

cob.!

[0082] The I-endonucleases can be classified as follows:

[0083] Class I: Two dodecapeptide motifs, 4 bp staggered cut with 3' OH overhangs, cut internal to recognition site

Subclass "I-SceI"	Other subclasses
I-SceI	I-SceII
I-SceIV	I-SceIII
I-CsmI	I-CeuI (only one dodecapeptide motif)
I-PanI	I-CreI (only one dodecapeptide motif)
	HO
	TFP1-408 (HO homolog)
	Endo SceI

[0084] Class II: GIY-(N_{10-11}) YIG motif, 2 bp staggered cut with 3' OH overhangs, cut external to recognition site:

[0085] I-TevI

[0086] Class III: no typical structural motifs, 4 bp staggered cut with 3' OH overhangs, cut internal to recognition site:

[0087] I-PpoI

[0088] Class IV: no typical structural motifs, 2 bp staggered cut with 3' OH overhangs, cut external to recognition site:

[0089] I-TevII

[0090] Class V: no typical structural motifs, 2 bp staggered cut with 5' OH overhangs:

[0091] I-TevIII.

[0092] 2. Nucleotide Probes Containing the I-SceI Gene of The Invention

[0093] The DNA sequence of the invention coding for the enzyme I-SceI can also be used as a probe for the detection of a nucleotide sequence in a biological material, such as tissue or body fluids. The probe can be labeled with an atom or inorganic radical, most commonly using a radionuclide, but also perhaps with a heavy metal. Radioactive labels include ^{32}P, ^{3}H, ^{14}C, or the like. Any radioactive label can be employed, which provides for an adequate signal and has sufficient half-life. Other labels include ligands that can serve as a specific binding member to a labeled antibody, fluorescers, chemiluminescers, enzymes, antibodies which can serve as a specific binding pair member for a labeled ligand, and the like. The choice of the label will be governed by the effect of the label on the rate of hybridization and binding of the probe to the DNA or RNA. It will be necessary that the label provide sufficient sensitivity to detect the amount of DNA or RNA available for hybridization.

[0094] When the nucleotide sequence of the invention is used as a probe for hybridizing to a gene, the nucleotide sequence is preferably affixed to a water insoluble solid, porous support, such as nitrocellulose paper. Hybridization can be carried out using labeled polynucleotides of the invention and conventional hybridization reagents. The particular hybridization technique is not essential to the invention.

[0095] The amount of labeled probe present in the hybridization solution will vary widely, depending upon the nature of the label, the amount of the labeled probe which can reasonably bind to the support, and the stringency of the hybridization. Generally, substantial excesses of the probe over stoichiometric will be employed to enhance the rate of binding of the probe to the fixed DNA.

[0096] Various degrees of stringency of hybridization can be employed. The more severe the conditions, the greater the complementarity that is required for hybridization between the probe and the polynucleotide for duplex formation. Severity can be controlled by temperature, probe concentration, probe length, ionic strength, time, and the like. Conveniently, the stringency of hybridization is varied by changing the polarity of the reactant solution. Temperatures to be employed can be empirically determined or determined from well known formulas developed for this purpose.

[0097] 3. Nucleotide Sequences Containing the Nucleotide Sequence Encoding I-SceI

[0098] This invention also relates to the DNA sequence of the invention encoding the enzyme I-SceI, wherein the nucleotide sequence is linked to other nucleic acids. The nucleic acid can be obtained from any source, for example, from plasmids, from cloned DNA or RNA, or from natural DNA or RNA from any source, including prokaryotic and eukaryotic organisms. DNA or RNA can be extracted from a biological material, such as biological fluids or tissue, by a variety of techniques including those described by Mania-

tis et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, New York (1982). The nucleic acid will generally be obtained from a bacteria, yeast, virus, or a higher organism, such as a plant or animal. The nucleic acid can be a fraction of a more complex mixture, such as a portion of a gene contained in whole human DNA or a portion of a nucleic acid sequence of a particular microorganism. The nucleic acid can be a fraction of a larger molecule or the nucleic acid can constitute an entire gene or assembly of genes. The DNA can be in a single-stranded or double-stranded form. If the fragment is in single-stranded form, it can be converted to double-stranded form using DNA polymerase according to conventional techniques.

[0099] The DNA sequence of the invention can be linked to a structural gene. As used herein, the term "structural gene" refers to a DNA sequence that encodes through its template or messenger mRNA a sequence of amino acids characteristic of a specific protein or polypeptide. The nucleotide sequence of the invention can function with an expression control sequence, that is, a DNA sequence that controls and regulates expression of the gene when operatively linked to the gene.

[0100] 4. Vectors Containing the Nucleotide Sequence of the Invention

[0101] This invention also relates to cloning and expression vectors containing the DNA sequence of the invention coding for the enzyme I-SceI.

[0102] More particularly, the DNA sequence encoding the enzyme can be ligated to a vehicle for cloning the sequence. The major steps involved in gene cloning comprise procedures for separating DNA containing the gene of interest from prokaryotes or eukaryotes, cutting the resulting DNA fragment and the DNA from a cloning vehicle at specific sites, mixing the two DNA fragments together, and ligating the fragments to yield a recombinant DNA molecule. The recombinant molecule can then be transferred into a host cell, and the cells allowed to replicate to produce identical cells containing clones of the original DNA sequence.

[0103] The vehicle employed in this invention can be any double-stranded DNA molecule capable of transporting the nucleotide sequence of the invention into a host cell and capable of replicating within the cell. More particularly, the vehicle must contain at least one DNA sequence that can act as the origin of replication in the host cell. In addition, the vehicle must contain two or more sites for insertion of the DNA sequence encoding the gene of the invention. These sites will ordinarily correspond to restriction enzyme sites at which cohesive ends can be formed, and which are complementary to the cohesive ends on the promoter sequence to be ligated to the vehicle. In general, this invention can be carried out with plasmid, bacteriophage, or cosmid vehicles having these characteristics.

[0104] The nucleotide sequence of the invention can have cohesive ends compatible with any combination of sites in the vehicle. Alternatively, the sequence can have one or more blunt ends that can be ligated to corresponding blunt ends in the cloning sites of the vehicle. The nucleotide sequence to be ligated can be further processed, if desired, by successive exonuclease deletion, such as with the enzyme Bal 31. In the event that the nucleotide sequence of the invention does not contain a desired combination of cohesive ends, the sequence can be modified by adding a linker, an adaptor, or homopolymer tailing.

[0105] It is preferred that plasmids used for cloning nucleotide sequences of the invention carry one or more genes responsible for a useful characteristic, such as a selectable marker, displayed by the host cell. In a preferred strategy, plasmids having genes for resistance to two different drugs are chosen. For example, insertion of the DNA sequence into a gene for an antibiotic inactivates the gene and destroys drug resistance. The second drug resistance gene is not affected when cells are transformed with the recombinants, and colonies containing the gene of interest can be selected by resistance to the second drug and susceptibility to the first drug. Preferred antibiotic markers are genes imparting chloramphenicol, ampicillin, or tetracycline resistance to the host cell.

[0106] A variety of restriction enzymes can be used to cut the vehicle. The identity of the restriction enzyme will generally depend upon the identity of the ends on the DNA sequence to be ligated and the restriction sites in the vehicle. The restriction enzyme is matched to the restriction sites in the vehicle, which in turn is matched to the ends on the nucleic acid fragment being ligated.

[0107] The ligation reaction can be set up using well known techniques and conventional reagents. Ligation is carried out with a DNA ligase that catalyzes the formation of phosphodiester bonds between adjacent 5'-phosphate and the free 3'-hydroxy groups in DNA duplexes. The DNA ligase can be derived from a variety of microorganisms. The preferred DNA ligases are enzymes from *E. coli* and bacteriophage T4. T4 DNA ligase can ligate DNA fragments with blunt or sticky ends, such as those generated by restriction enzyme digestion. *E. coli* DNA ligase can be used to catalyze the formation of phosphodiester bonds between the termini of duplex DNA molecules containing cohesive ends.

[0108] Cloning can be carried out in prokaryotic or eukaryotic cells. The host for replicating the cloning vehicle will of course be one that is compatible with the vehicle and in which the vehicle can replicate. When a plasmid is employed, the plasmid can be derived from bacteria or some other organism or the plasmid can be synthetically prepared. The plasmid can replicate independently of the host cell chromosome or an integrative plasmid (episome) can be employed. The plasmid can make use of the DNA replicative enzymes of the host cell in order to replicate or the plasmid can carry genes that code for the enzymes required for plasmid replication. A number of different plasmids can be employed in practicing this invention.

[0109] The DNA sequence of the invention encoding the enzyme I-SceI can also be ligated to a vehicle to form an expression vector. The vehicle employed in this case is one in which it is possible to express the gene operatively linked to a promoter in an appropriate host cell. It is preferable to employ a vehicle known for use in expressing genes in *E. coli,* yeast, or mammalian cells. These vehicles include, for example, the following *E. coli* expression vectors:

[0110] pSCM525, which is an *E. coli* expression vector derived from pUC12 by insertion of a tac promoter and the synthetic gene for I-SceI. Expression is induced by IPTG.

[0111] pGEXω6, which is an *E. coli* expression vector derived from pGEX in which the synthetic gene from pSCM525 for I-SceI is fused with the glutathione S transferase gene, producing a hybrid protein. The hybrid protein possesses the endonuclease activity.

[0112] pDIC73, which is an *E. coli* expression vector derived from pET-3C by insertion of the synthetic gene for I-SceI (NdeI—BamHI fragment of pSCM525) under T7 promoter control. This vector is used in strain BL21 (DE3) which expresses the T7 RNA polymerase under IPTG induction.

[0113] pSCM351, which is an *E. coli* expression vector derived from pUR291 in which the synthetic gene for I-SceI is fused with the Lac Z gene, producing a hybrid protein.

[0114] pSCM353, which is an *E. coli* expression vector derived from PEX1 in which the synthetic gene for I-SceI is fused with the Cro/Lac Z gene, producing a hybrid protein.

[0115] Examples of yeast expression vectors are:

[0116] pPEX7, which is a yeast expression vector derived from pRP51-Bam O (a LEU2d derivative of pLG-SD5) by insertion of the synthetic gene under the control of the galactose promoter. Expression is induced by galactose.

[0117] pPEX408, which is a yeast expression vector derived from pLG-SD5 by insertion of the synthetic gene under the control of the galactose promoter. Expression is induced by galactose.

[0118] Several yeast expression vectors are depicted in **FIG. 7**.

[0119] Typical mammalian expression vectors are:

[0120] pRSV I-SceI, which is a pRSV derivative in which the synthetic gene (BamHI—PstI fragment from pSCM525) is under the control of the LTR promoter of Rous Sarcoma Virus. This expression vector is depicted in **FIG. 8**.

[0121] Vectors for expression in Chinese Hamster Ovary (CHO) cells can also be employed.

[0122] 5. Cells Transformed with Vectors of the Invention

[0123] The vectors of the invention can be inserted into host organisms using conventional techniques. For example, the vectors can be inserted by transformation, transfection, electroporation, microinjection, or by means of liposomes (lipofection).

[0124] Cloning can be carried out in prokaryotic or eukaryotic cells. The host for replicating the cloning vehicle will of course be one that is compatible with the vehicle and in which the vehicle can replicate. Cloning is preferably carried out in bacterial or yeast cells, although cells of fungal, animal, and plant origin can also be employed. The preferred host cells for conducting cloning work are bacterial cells, such as *E. coli*. The use of *E. coli* cells is particularly preferred because most cloning vehicles, such as bacterial plasmids and bacteriophages, replicate in these cells.

[0125] In a preferred embodiment of this invention, an expression vector containing the DNA sequence encoding the nucleotide sequence of the invention operatively linked to a promoter is inserted into a mammalian cell using conventional techniques.

Application of I-SceI for Large Scale Mapping

[0126] 1. Occurrence of Natural Sites in Various Genomes

[0127] Using the purified I-SceI enzyme, the occurrence of natural or degenerate sites has been examined on the complete genomes of several species. No natural site was found in *Saccharomyces cerevisiae, Bacillus anthracis, Borrelia burgdorferi, Leptospira biflexa* and *L. interrogans*. One degenerate site was found on T7 phage DNA.

[0128] 2. Insertion of Artificial Sites

[0129] Given the absence of natural I-SceI sites, artificial sites can be introduced by transformation or transfection. Two cases need to be distinguished: site-directed integration by homologous recombination and random integration by non-homologous recombination, transposon movement or retroviral infection. The first is easy in the case of yeast and a few bacterial species, more difficult for higher eucaryotes. The second is possible in all systems.

[0130] 3. Insertion Vectors

[0131] Two types can be distinguished:

[0132] 1—Site specific cassettes that introduce the I-SceI site together with a selectable marker.

[0133] For yeast: all are pAF100 derivatives (Thierry et al. (1990) YEAST 6:521-534) containing the following marker genes:

pAF101:	URA3 (inserted in the HindIII site)
pAF103:	Neo^R (inserted in BglII site)
pAF104:	HIS3 (inserted in BglII site)
pAF105:	Kan^R (inserted in BglII site)
pAF106:	Kan^R (inserted in BglII site)
pAF107:	LYS2 (inserted between HindIII and EcoR V)

[0134] A restriction map of the plasmid pAF100 is shown in **FIG. 9**. The nucleotide sequence and restriction sites of regions of plasmid pAF100 are shown in **FIGS. 10A and 10B**. Many transgenic yeast strains with the I-SceI site at various and known places along chromosomes are available.

[0135] 2—Vectors derived from transposable elements or retroviruses.

[0136] For *E. coli* and other bacteria: mini Tn5 derivatives containing the I-SceI site and

[0137] pTSm ω Str^R

[0138] pTKm ω Kan^R (See **FIG. 11**)

[0139] pTTc ω Tet^R

[0140] For yeast: pTyω6 is a pD123 derivative in which the I-SceI site has been inserted in the LTR of the Ty element. (**FIG. 12**)

[0141] For mammalian cells:

[0142] pMLV LTR SAPLZ: containing the I-SceI site in the LTR of MLV and Phleo-LacZ (**FIG. 13**). This vector is first grown in ψ2 cells (3T3 derivative, from R. Mulligan). Two transgenic cell lines with the I-SceI site at undetermined locations in the genome are available: 1009 (pluripotent nerve cells, J. F. Nicolas) and D3 (ES cells able to generate transgenic animals).

[0143] 4. The Nested Chromosomal Fragmentation Strategy

[0144] The nested chromosomal fragmentation strategy for genetically mapping a eukaryotic genome exploits the unique properties of the restriction endonuclease I-SceI, such as an 18 bp long recognition site. The absence of natural I-SceI recognition sites in most eukaryotic genomes is also exploited in this mapping strategy.

[0145] First, one or more I-SceI recognition sites are artificially inserted at various positions in a genome, by homologous recombination using specific cassettes containing selectable markers or by random insertion, as discussed supra. The genome of the resulting transgenic strain is then cleaved completely at the artificially inserted I-SceI site(s) upon incubation with the I-SceI restriction enzyme. The cleavage produces nested chromosomal fragments.

[0146] The chromosomal fragments are then purified and separated by pulsed field gel (PFG) electrophoresis, allowing one to "map" the position of the inserted site in the chromosome. If total DNA is cleaved with the restriction enzyme, each artificially introduced I-SceI site provides a unique "molecular milestone" in the genome. Thus, a set of transgenic strains, each carrying a single I-SceI site, can be created which defines physical genomic intervals between the milestones. Consequently, an entire genome, a chromosome or any segment of interest can be mapped using artificially introduced I-SceI restriction sites.

[0147] The nested chromosomal fragments may be transferred to a solid membrane and hybridized to a labelled probe containing DNA complementary to the DNA of the fragments. Based on the hybridization banding patterns that are observed, the eukaryotic genome may be mapped. The set of transgenic strains with appropriate "milestones" is used as a reference to map any new gene or clone by direct hybridization.

EXAMPLE 1

[0148] Application of the Nested Chromosomal Fragmentation Strategy to the Mapping of Yeast Chromosome XI

[0149] This strategy has been applied to the mapping of yeast chromosome XI of *Saccharamyces cerevisiae*. The I-SceI site was inserted at 7 different locations along chromosome XI of the diploid strain FY1679, hence defining eight physical intervals in that chromosome. Sites were inserted from a URA3-1-I-SceI cassette by homologous recombination. Two sites were inserted within genetically defined genes, TIF1 and FAS1, the others were inserted at unknown positions in the chromosome from five non-overlapping cosmids of our library, taken at random. Agarose embedded DNA of each of the seven transgenic strains was then digested with I-SceI and analyzed by pulsed field gel electrophoresis (**FIG. 14A**). The position of the I-SceI site of each transgenic strain in chromosome XI is first deduced from the fragment sizes without consideration of the left/right orientation of the fragments. Orientation was determined as follows. The most telomere proximal I-SceI site from this set of strains is in the transgenic E40 because the 50 kb fragment is the shortest of all fragments (**FIG. 15A**). Therefore, the cosmid clone pUKGO40, which was used to insert the I-SceI site in the transgenic E40, is now used as a probe against all chromosome fragments (**FIG. 14B**). As expected, pUKG040 lights up the two fragments from strain E40 (50 kb and 630 kb, respectively). The large fragment is close to the entire chromosome XI and shows a weak hybridization signal due to the fact that the insert of pUKG040, which is 38 kb long, contains less than 4 kb within the large chromosome fragment. Note that the entire chromosome XI remains visible after I-SceI digestion, due to the fact that the transgenic strains are diploids in which the I-SceI site is inserted in only one of the two homologs. Now, the pUKG040 probe hybridizes to only one fragment of all other transgenic strains allowing unambiguous left/right orientation of I-SceI sites (See **FIG. 15B**). No significant cross hybridization between the cosmid vector and the chromosome subfragment containing the I-SceI site insertion vector is visible. Transgenic strains can now be ordered such that I-SceI sites are located at increasing distances from the hybridizing end of the chromosome (**FIG. 15C**) and the I-SceI map can be deduced (**FIG. 15D**). Precision of the mapping depends upon PFGE resolution and optimal calibration. Note that actual left/right orientation of the chromosome with respect to the genetic map is not known at this step. To help visualize our strategy and to obtain more precise measurements of the interval sizes between I-SceI sites between I-SceI, a new pulsed field gel electrophoresis with the same transgenic strains now placed in order was made (**FIG. 16**). After transfer, the fragments were hybridized successively with cosmids pUKG040 and pUKG066 which light up, respectively, all fragments from the opposite ends of the chromosome (clone pUKG066 defines the right end of the chromosome as defined from the genetic map because it contains the SIR1 gene. A regular stepwise progression of chromosome fragment sizes is observed. Note some cross hybridization between the probe pUKG066 and chromosome III, probably due to some repetitive DNA sequences.

[0150] All chromosome fragments, taken together, now define physical intervals as indicated in **FIG. 15*d***. The I-SceI map obtained has an 80 kb average resolution.

EXAMPLE 2

[0151] Application of the Nested Chromosomal Fragmentation Strategy to the Mapping of Yeast Artificial Chromosome (YAC) Clones

[0152] This strategy can be applied to YAC mapping with two possibilities.

[0153] 1—insertion of the I-SceI site within the gene of interest using homologous recombination in yeast. This permits mapping of that gene in the YAC insert by I-SceI digestion in vitro. This has been done and works.

[0154] 2—random integration of I-SceI sites along the YAC insert by homologous recombination in yeast using highly repetitive sequences (e.g., B2 in

0 (use "zero" instead of capital letter "O")

mouse or Alu in human). Transgenic strains are then used as described in ref. P1 to sort libraries or map genes.

[0155] The procedure has now been extended to YAC containing 450 kb of Mouse DNA. To this end, a repeated sequence of mouse DNA (called B2) has been inserted in a plasmid containing the I-SceI site and a selectable yeast marker (LYS2). Transformation of the yeast cells containing the recombinant YAC with the plasmid linearized within the B2 sequence resulted in the integration of the I-SceI site at five different locations distributed along the mouse DNA insert. Cleavage at the inserted I-SceI sites using the enzyme has been successful, producing nested fragments that can be purified after electrophoresis. Subsequent steps of the protocol exactly parallels the procedure described in Example 1.

EXAMPLE 3

[0156] Application of Nested Chromosomal Fragments to the Direct Sorting of Cosmid Libraries

[0157] The nested, chromosomal fragments can be purified from preparative PFG and used as probes against clones from a chromosome X1 specific sublibrary. This sublibrary is composed of 138 cosmid clones (corresponding to eight times coverage) which have been previously sorted from our complete yeast genomic libraries by colony hybridization with PFG purified chromosome X1. This collection of unordered clones has been sequentially hybridized with chromosome fragments taken in order of increasing sizes from the left end of the chromosome. Localization of each cosmid clone on the I-SceI map could be unambiguously determined from such hybridizations. To further verify the results and to provide a more precise map, a subset of all cosmid clones, now placed in order, have been digested with EcoRI, electrophoresed and hybridized with the nested series of chromosome fragments in order of increasing sizes from the left end of the chromosome. Results are given in **FIG. 17**.

[0158] For a given probe, two cases can be distinguished: cosmid clones in which all EcoRI fragments hybridize with the probe and cosmid clones in which only some of the EcoRI fragments hybridize (i.e., compare pEKG100 to pEKG098 in FIG. 17*b*). The first category corresponds to clones in which the insert is entirely included in one of the two chromosome fragments, the second to clones in which the insert overlaps an I-SceI site. Note that, for clones of the pEKG series, the EcoRI fragment of 8 kb is entirely composed of vector sequences (pWE15) that do not hybridize with the chromosome fragments. In the case where the chromosome fragment possesses the integration vector, a weak cross hybridization with the cosmid is observed (**FIG. 17***e*).

[0159] Examination of **FIG. 17** shows that the cosmid clones can unambiguously be ordered with respect to the I-SceI map (**FIG. 13E**), each clone falling either in a defined interval or across an I-SceI site. In addition, clones from the second category allow us to place some EcoRI fragments on the I-SceI maps, while others remain unordered. The complete set of chromosome XI-specific cosmid clones, covering altogether eight times the equivalent of the chromosome, has been sorted with respect to the I-SceI map, as shown in **FIG. 18**.

[0160] 5. Partial Restriction Mapping Using I-SceI

[0161] In this embodiment, complete digestion of the DNA at the artificially inserted I-SceI site is followed by partial digestion with bacterial restriction endonucleases of choice. The restriction fragments are then separated by electrophoresis and blotted. Indirect end labelling is accomplished using left or right I-Sce half sites. This technique has been successful with yeast chromosomes and should be applicable without difficulty for YAC.

[0162] Partial restriction mapping has been done on yeast DNA and on mammalian cell DNA using the commercial enzyme I-SceI. DNA from cells containing an artificially inserted I-SceI site is first cleaved to completion by I-SceI. The DNA is then treated under partial cleavage conditions with bacterial restriction endonucleases of interest (e.g., BamHI) and electrophoresed along with size calibration markers. The DNA is transferred to a membrane and hybridized successively using the short sequences flanking the I-SceI sites on either side (these sequences are known because they are part of the original insertion vector that was used to introduce the I-SceI site). Autoradiography (or other equivalent detection system using non radioactive probes) permit the visualization of ladders, which directly represent the succession of the bacterial restriction endonuclease sites from the I-SceI site. The size of each band of the ladder is used to calculate the physical distance between the successive bacterial restriction endonuclease sites.

Application of I-SceI for In Vivo Site Directed Recombination

[0163] 1. Expression of I-SceI in Yeast

[0164] The synthetic I-SceI gene has been placed under the control of a galactose inducible promoter on multicopy plasmids pPEX7 and pPEX408. Expression is correct and induces effects on site as indicated below. A transgenic yeast with the I-SceI synthetic gene inserted in a chromosome under the control of an inducible promoter can be constructed.

[0165] 2. Effects of Site Specific Double Strand Breaks in Yeast (Refs. 18 and P4)

[0166] Effects on plasmid-borne I-SceI sites:

[0167] Intramolecular effects are described in detail in Ref. 18. Intermolecular (plasmid to chromosome) recombination can be predicted.

Effects on Chromosome Integrated I-SceI Sites

[0168] In a haploid cell, a single break within a chromosome at an artificial I-SceI site results in cell division arrest followed by death (only a few % of survival). Presence of an intact sequence homologous to the cut site results in repair and 100% cell survival. In a diploid cell, a single break within a chromosome at an artificial I-SceI site results in repair using the chromosome homolog and 100% cell survival. In both cases, repair of the induced double strand break results in loss of heterozygosity with deletion of the non homologous sequences flanking the cut and insertion of the non homologous sequences from the donor DNA molecule.

[0197] 7. B. Dujon, G. Cottarel, L. Colleaux, M. Betermier, A. Jacquier, L. D'Auriol, F. Galibert, Mechanism of integration of an intron within a mitochondrial gene: a double strand break and the transposase function of an intron encoded protein as revealed by in vivo and in vitro assays. In Achievements and perspectives of Mitochondrial Research. Vol. II, Biogenesis, E. Quagliariello et al. Eds. Elsevier, Amsterdam (1985) pages 215-225.

[0198] 8. L. Colleaux, L. D'Auriol, M. Betermier, G. Cottarel, A.

[0199] Jacquier, F. Galibert, and B. Dujon, A universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *Escherichia coli* as a specific double strand endonuclease. Cell (1986) 44, 521-533.

[0200] 9. B. Dujon, L. Colleaux, A. Jacquier, F. Michel and C.

[0201] Monteilhet, Mitochondrial introns as mobile genetic elements: the role of intron-encoded proteins. In "Extrachromosomal elements in lower eucaryotes", Reed B et al. Eds. (1986) Plenum Pub. Corp. 5-27.

[0202] 10. F. Michel and B. Dujon, Genetic Exchanges between Bacteriophage T4 and Filamentous Fungi? Cell (1986) 46, 323.

[0203] 11. L. Colleaux, L. D'Auriol, F. Galibert and B. Dujon, Recognition and cleavage site of the intron encoded omega transposase. PNAS (1988), 85, 6022-6026.

[0204] 12. B. Dujon, Group I introns as mobile genetic elements, facts and mechanistic speculations: A Review. Gene (1989), 82, 91-114.

[0205] 13. B. Dujon, M. Belfort, R. A. Butow, C.Jacq, C. Lemieux, P. S. Perlman, V. M. Vogt, Mobile introns: definition of terms and recommended nomenclature. Gene (1989), 82, 115-118.

[0206] 14. C. Monteilhet, A. Perrin, A. Thierry, L. Colleaux, B. Dujon, Purification and Characterization of the in vitro activity of I-SceI, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acid Research (1990), 18, 1407-1413.

[0207] 15. L. Colleaux, M-R. Michel-Wolwertz, R. F. Matagne, B. Dujon—The apocytochrome b gene of *Chlamydomonas smithii* contains a mobile intron related to both Saccharomyces and Neurospora introns. Mol. Gen. Genet. (1990) 223, 288-296.

[0208] 16. B. Dujon Des introns autonomes et mobiles. Annales de l'Institut Pasteur/Actualites (1990) 1. 181-194.

[0209] 17. A. Thierry, A. Perrin, J. Boyer, C. Fairhead, B. Dujon, B. Frey, G. Schmitz. Cleavage of yeast and bacteriophage 17 genomes at a single site using the rare cutter endonuclease I-Sce. I Nuc. Ac. Res. (1991) 19, 189-190.

[0210] 18. A. Plessis, A. Perrin, J. E. Haber, B. Dujon, Site specific recombination determined by I-SceI, a mitochondrial intron-encoded endonuclease expressed in the yeast nucleus. GENETICS (1992) 130, 451-460.

Abstracts

[0211] A1. A. Jacquier, B. Dujon. Intron recombinational insertion at the DNA level: Nature of a specific receptor site and direct role of an intron encoded protein. Cold Spring Harbor Symposium 1984.

[0212] A2. I. Colleaux, L. D'Auriol, M. Demariaux, B. Dujon, F. Galibert, and A. Jacquier, Construction of a universal code equivalent from a mitochondrial intron encoded transposase gene using oligonucleotide directed multiple mutagenesis. Colloque International de DNRS "oligonucleotids et Genetique Moleculaire" Aussois (Savoie) Jan. 8-12, 1985.

[0213] A3. L. Colleaux, D'Auriol, M. Demariaux, B. Dujon, F. Galibert, and A. Jacquier, Expression in *E. coli* of a universal code equivalent of a yeast mitochondrial intron reading frame involved in the integration of an intron within a gene. Cold Spring Harbor Meeting on "Molecular Biology of Yeast", Aug. 13-19, 1985.

[0214] A4. B. Dujon, G. Cottarel, L. Colleaux, M. Demariaux, A. Jacquier, L. D'Auriol, and F. Galibert, Mechanism of integration of an intron within a mitochondrial gene: a double strand break and the "transposase" function of an intron encoded protein as revealed by in vivo and in vitro assays. International symposium on "Achievements and Perspectives in Mitochondrial Research", Selva de Fasono (Brindisi, Italy) Sep. 26, 1985.

[0215] 5. L. Colleaux, G. Cottarel, M. Betermier, A. Jacquier, B. Dujon, L. D'auriol, and F. Galibert, Mise en evidence de l'activite endonuclease double brin d'unc protein codee par un intron mitochondrial de levure. Forum sur la Biologie Moleculaire de la levure, Bonbannes, France Oct. 2-4, 1985.

[0216] A6. B. Dujon, L. Colleaux, F. Michel and A. Jacquier, Mitochondrial introns as mobile genetic elements. In "Extrachromosomal elements in lower eucaryotes", Urbana, Ill., Jun. 1-5, 1986.

[0217] A7. L. Colleaux and B. Dujon, Activity of a mitochondrial intron encoded transposase. Yeast Genetics and Molecular Biology Meeting, Urbana, Ill. Jun. 3-6, 1986.

[0218] A8. L. Colleaux and B. Dujon, The role of a mitochondrial intron encoded protein. XIIIth International Conference on Yeast Genetics and Molecular Biology, Banff, Alberta (Canada) Aug. 31-Sep. 5, 1986.

[0219] A9. L. Colleaux, L. D'Aurio, F. Galibert and and B. Dujon, Recognition and cleavage specificity of an intron encoded transposase. 1987 Meeting on Yeast Genetics and Molecular Biology. San Francisco, Calif. Jun. 16-21, 1987.

[0220] A10. A. Perrin, C. Monteilhet, L. Colleaux and B. Dujon, Biochemical activity of an intron encoded transposase of *Saccharomyces cerevisiae.*

Cold Spring Harbor Meeting on "Molecular Biology of Mitochondria and chloroplasts" Aug. 25-30, 1987 Cold Spring Harbor, N.Y.

No ¶ break here

[0221] A11. B. Dujon, A. Jacquier, L. Colleaux, C. Monteilhet, A.

[0222] Perrin, "Les Introns autoepissables et leurs proteins" Colloque "Biologie Moleculaire de la levure: expression genetique chez Saccharomyces" organise par la Societe francaise de Microbiologie Jan. 18, 1988 Institut Pasteur, Paris.

[0223] A12. L. Colleaux, L. D'Auriol, C. Monteilhet, F. Galibert and B. Dujon, Characterization of the biochemical activity of an intron encoded transposase. 14th International Conference on Yeast Genetics and Molecular Biology. Espoo, Finland, Aug. 7-13, 1988.

[0224] A13. B. Dujon, A goup I intron as a mobile genetic element, Albany Conference sur "RNA: catalysis, splicing, evolution", Albany, N.Y., Sep. 22-25, 1988.

[0225] A14. B. Dujon, L. Colleaux, C. Monteilhet, A. Perrin, L. D'Auriol, F. Galibert, Group I introns as mobile genetic elements: the role of intron encoded proteins and the nature of the target site. 14th Annual EMBO Symposium "Organelle genomes and the nucleus" Heidelberg, Sep. 26-29, 1988.

[0226] A15. L. Colleaux, R. Matagne, B. Dujon, A new mobile mitochondrial intron provides evidence for genetic exchange between Neurospora and Chlamydomonas species. Cold Spring Harbor, May 1989.

[0227] A16. L. Colleaux, M. R. Michel-Wolwertz, R. F. Matagne, B. Dujon, The apoxytochrome b gene of *Chlamydomonas smithii* contains a mobile intron related to both Saccharomyces and Neurospora introns. Fourth International Conference on Cell and Molecular Biology of Chlamydomonas.

No ¶ break here

[0228] Madison, Wis., April 1990.

[0229] A17. B. Dujon, L. Colleaux, E. Luzi, C. Monteilhet, A. Perrin, A. Plessis, I. Stroke, A. Thierry, Mobile Introns, EMBO Workshop on "Molecular Mechanisms of transposition and its control, Roscoff (France) June 1990.

[0230] A18. A. Perrin, C. Monteilhet, A. Thierry, E. Luzi, I. Stroke, L. Colleaux, B. Dujon. I-SceI, a novel double strand site specific endonuclease, encoded by a mobile group I intron in Yeast. Workshop on "RecA and Related Proteins" Sacly, France Sep. 17-21, 1990.

[0231] A19. A. Plessis, A. Perrin, B. Dujon, Site specific recombination induced by double strand endonucleases, HO and I-SceI in yeast. Workshop on "RecA and Related Proteins" Saclay, France Sep. 17-21, 1990.

[0232] A20. B. Dujon, The genetic propagation of introns 20th FEBS Meeting, Budapest, Hungary, August 1990.

[0233] A21. E. Luzi, B. Dujon, Analysis of the intron encoded site specific endonuclease I-SceI by mutagenesis, Third European Congress on Cell Biology, Florence, Italy, September 1990.

[0234] A22. B. Dujon, Self splicing introns as contagious genetic elements. Journees Franco-Beiges de Pont a Mousson. October 1990.

[0235] A23. B. Frey, H. Dubler, G. Schmitz, A. Thierry, A. Perrin, J. Boyer, C. Fairhead, B. Dujon, Specific cleavage of the yeast genome at a single site using the rare cutter endonuclease I-SceI Human Genome, Frankfurt, Germany, November 1990.

[0236] A24. B. Dujon, A. Perrin, I. Stroke, E. Luzi, L. Colleaux, A.

No ¶ break here

[0237] Plessis, A. Thierry, The genetic mobility of group I introns at the DNA level. Keystone Symposia Meeting on "Molecular Evolution of Introns and Other RNA elements", Taos, N. Mex., Feb. 2-8, 1991.

[0238] A25. B. Dujon, J. Boyer, C. Fairhead, A. Perrin, A Thierry, Cartographie chez la levure. Reunion "Strategies d'etablissement des cartes geniques" Toulouse Mai 30-31. 1991.

[0239] A26. B. Dujon, A. Thierry, Nested chromosomal fragmentation using the meganuclease I-SceI: a new method for the rapid mapping of the yeast genome. Elounda, Crete 15-17 Mail 1991.

[0240] A27. A. Thierry, L. Gaillon, F. Galibert, B. Dujon. The chromosome XI library: what has been accomplished, what is left. Brugge meeting Sep. 22-24, 1991.

[0241] A28. B. Dujon, A. Thierry, Nested chromosomal fragmentation using the meganuclease I-SceI: a new method for the rapid physical mapping of the eukaryotic genomes. Cold Spring Harbor May 6-10, 1992.

[0242] A29. A. Thierry, L. Gaillon, F. Galibert, B. Dujon. Yeast chromosome XI: construction of a cosmid contig. a high resolution map and sequencing progress. Cold Spring Harbor May 6-10, 1992.

In Preparation

[0243] P1. A. Thierry and B. Dujon, Nested Chromosomal Fragmentation Using the Meganuclease I-SceI: Application to the physical mapping of a yeast chromosome and the direct sorting of cosmid libraries. Probably Submission to GENOMICS or EMBO J.

[0244] P2. A. Thierry, L. Colleaux and B. Dujon: Construction and Expression of a synthetic gene coding for the meganuclease I-SceI. Possible submission: NAR, EMBO J.

[0245] P3. I. Stroke, V. Pelicic and B. Dujon: The evolutionarily conserved dodecapeptide motifs of intron-encoded I-SceI are essential for endonuclease function. Submission to EMBO J.

[0246] P4. C. Fairhead and B. Dujon: Consequences of a double strand break induced in vivo in yeast at

-continued

```
CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA AAAACGCCAG CAACGCGGCC    180

TTTTTACGGT TCCTGGCCTT TTGCTGGCCT TTTGCTCACA TGTTCTTTCC TGCGTTATCC    240

CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCAGC    300

CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCC AATACGCAAA    360

CCGCCTCTCC CCGCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG GTTTCCCGAC    420

TGGAAAGCGG GCAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC    480

CAGGCTTTAC ACTTTATGCT TCCGGCTCGT ATGTTGTGTG GAATTGTGAG CGGATAACAA    540

TTTCACACAG GAAACAGCTA TGACCATGAT TACGAATTCT CATGTTTGAC AGCTTATCAT    600

CGATAAGCTT TAATGCGGTA GTTTATCACA GTTAAATTGC TAACGCAGTC AGGCACCGTG    660

TATGAAATCT AACAATGCGC TCATCGTCAT CCTCGGCACC GTCACCCTGG ATGCTGTAGG    720

CATAGGCTTG GTTATGCCGG TACTGCCGGG CCTCTTGCGG GATATCCGCC TGATGCGTGA    780

ACGTGACGGA CGTAACCACC GCGACATGTG TGTGCTGTTC CGCTGGGCAT GCCAGGACAA    840

CTTCTGGTCC GGTAACGTGC TGAGCCCGGC CAAGCTTACT CCCCATCCCC CTGTTGACAA    900

TTAATCATCG GCTCGTATAA TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA    960

CAGGATCCAT GCATATGAAA AACATCAAAA AAAACCAGGT AATGAACCTG GGTCCGAACT   1020

CTAAACTGCT GAAAGAATAC AAATCCCAGC TGATCGAACT GAACATCGAA CAGTTCGAAG   1080

CAGGTATCGG TCTGATCCTG GGTGATGCTT ACATCCGTTC TCGTGATGAA GGTAAAACCT   1140

ACTGTATGCA GTTCGAGTGG AAAAACAAAG CATACATGGA CCACGTATGT CTGCTGTACG   1200

ATCAGTGGGT ACTGTCCCCG CCGCACAAAA AAGAACGTGT TAACCACCTG GGTAACCTGG   1260

TAATCACCTG GGGCGCCCAG ACTTTCAAAC ACCAAGCTTT CAACAAACTG GCTAACCTGT   1320

TCATCGTTAA CAACAAAAAA ACCATCCCGA ACAACCTGGT TGAAAACTAC CTGACCCCGA   1380

TGTCTCTGGC ATACTGGTTC ATGGATGATG GTGGTAAATG GGATTACAAC AAAAACTCTA   1440

CCAACAAATC GATCGTACTG AACACCCAGT CTTTCACTTT CGAAGAAGTA GAATACCTGG   1500

TTAAGGGTCT GCGTAACAAA TTCCAACTGA ACTGTTACGT AAAAATCAAC AAAAACAAAC   1560

CGATCATCTA CATCGATTCT ATGTCTTACC TGATCTTCTA CAACCTGATC AAACCGTAAT   1620

TCATCCCCCA GATGATGTAC AAACTGCCGA ACACTATCTC CTCCGAAACT TTCCTGAAAT   1680

AATAAGTCGA CCTGCAGCCC AAGCTTGGCA CTGGCCGTCG TTTTACAACG TCGTGACT     1738
```

CC

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
Met Leu Val Arg Gly Ala Glu Pro Met Glu Lys Arg Gln Gln Arg Gly
1               5                  10                  15

Leu Phe Thr Val Pro Gly Leu Leu Leu Ala Phe Cys Ser His Val Leu
            20                  25                  30

Ser Cys Val Ile Pro
        35
```

-continued

Ser Cys Val Ile Pro
35

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Gln Leu Ala Arg Gln Val Ser Arg Leu Glu Ser Gly Gln
1 5 10

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Leu Pro Ala Arg Met Leu Cys Gly Ile Val Ser Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Thr Met Ile Thr Asn Ser His Val
1 5

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 80 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Lys Ser Asn Asn Ala Leu Ile Val Ile Leu Gly Thr Val Thr Leu
1 5 10 15

Asp Ala Val Gly Ile Gly Leu Val Met Pro Val Leu Pro Gly Leu Leu
20 25 30

Arg Asp Ile Arg Leu Met Arg Glu Arg Asp Gly Arg Asn His Arg Asp
35 40 45

Met Cys Val Leu Phe Arg Trp Ala Cys Gln Asp Asn Phe Trp Ser Gly
50 55 60

Asn Val Leu Ser Pro Ala Lys Leu Thr Pro His Pro Pro Val Asp Asn
65 70 75 80

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

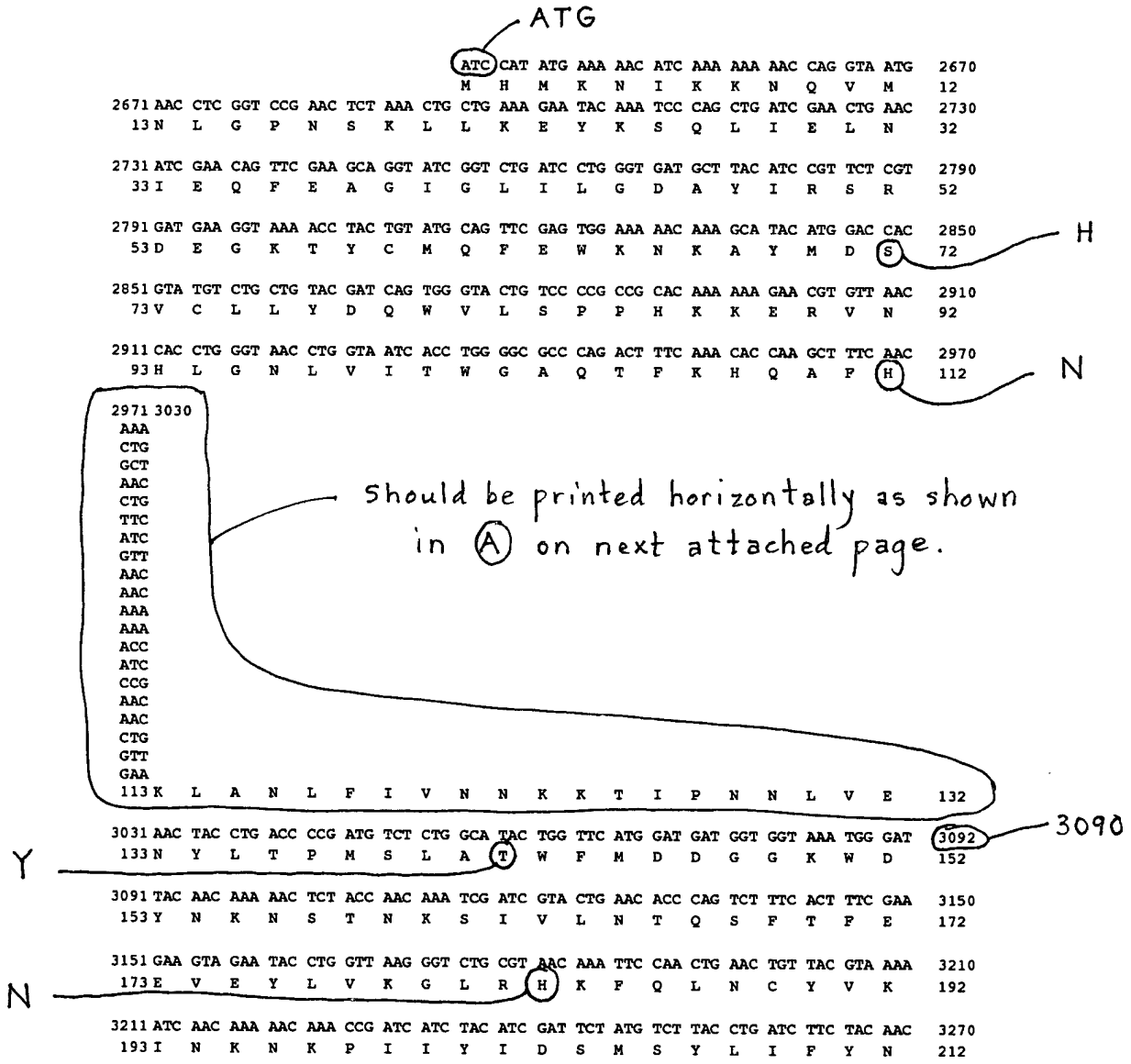
(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

ATCCCTATTG TCCCATTA 18

What is claimed is:

1. An isolated DNA encoding the enzyme I-SceI, wherein the DNA has the nucleotide squence:




```
                                  ATC CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG   2670
                                    M   H   M   K   N   I   K   K   N   Q   V   M    12
2671 AAC CTC GGT CCG AAC TCT AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC 2730
  13 N   L   G   P   N   S   K   L   L   K   E   Y   K   S   Q   L   I   E   L   N   32

2731 ATC GAA CAG TTC GAA GCA GGT ATC GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT 2790
  33 I   E   Q   F   E   A   G   I   G   L   I   L   G   D   A   Y   I   R   S   R   52

2791 GAT GAA GGT AAA ACC TAC TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC 2850
  53 D   E   G   K   T   Y   C   M   Q   F   E   W   K   N   K   A   Y   M   D   S   72

2851 GTA TGT CTG CTG TAC GAT CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC 2910
  73 V   C   L   L   Y   D   Q   W   V   L   S   P   P   H   K   K   E   R   V   N   92

2911 CAC CTG GGT AAC CTG GTA ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC 2970
  93 H   L   G   N   L   V   I   T   W   G   A   Q   T   F   K   H   Q   A   F   H  112

2971 3030
 AAA
 CTG
 GCT
 AAC
 CTG
 TTC
 ATC
 GTT
 AAC
 AAC
 AAA
 AAA
 ACC
 ATC
 CCG
 AAC
 AAC
 CTG
 GTT
 GAA
 113 K   L   A   N   L   F   I   V   N   N   K   K   T   I   P   N   N   L   V   E  132

3031 AAC TAC CTG ACC CCG ATG TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT 3092
 133 N   Y   L   T   P   M   S   L   A   T   W   F   M   D   D   G   G   K   W   D  152

3091 TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA 3150
 153 Y   N   K   N   S   T   N   K   S   I   V   L   N   T   Q   S   F   T   F   E  172

3151 GAA GTA GAA TAC CTG GTT AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC GTA AAA 3210
 173 E   V   E   Y   L   V   K   G   L   R   H   K   F   Q   L   N   C   Y   V   K  192

3211 ATC AAC AAA AAC AAA CCG ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC 3270
 193 I   N   K   N   K   P   I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N  212
```

What is claimed is:

1. An isolated DNA encoding the enzyme I-*Sce*I, wherein the DNA has the nucleotide squence:

```
                                                ATG CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG 2670
                                                M   H   M   K   N   I   K   K   N   Q   V   M   12

2671 AAC CTC GGT CCG AAC TCT AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC 2730
  13 N   L   G   P   N   S   K   L   L   K   E   Y   K   S   Q   L   I   E   L   N   32

2731 ATC GAA CAG TTC GAA GCA GGT ATC GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT 2790
  33 I   E   Q   F   E   A   G   I   G   L   I   L   G   D   A   Y   I   R   S   R   52

2791 GAT GAA GGT AAA ACC TAC TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC 2850
  53 D   E   G   K   T   Y   C   M   Q   F   E   W   K   N   K   A   Y   M   D   H   72

2851 GTA TGT CTG CTG TAC GAT CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC 2910
  73 V   C   L   L   Y   D   Q   W   V   L   S   P   P   H   K   K   E   R   V   N   92

2911 CAC CTG GGT AAC CTG GTA ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC 2970
  93 H   L   G   N   L   V   I   T   W   G   A   Q   T   F   K   H   Q   A   F   N  112

(A) =
2971 AAA CTG GCT AAC CTG TTC ATC GTT AAC AAC AAA AAA ACC ATC CCG AAC AAC CTG GTT GAA 3030
 113 K   L   A   N   L   F   I   V   N   N   K   K   T   I   P   N   N   L   V   E  132

3031 AAC TAC CTG ACC CCG ATG TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT 3090
 133 N   Y   L   T   P   M   S   L   A   Y   W   F   M   D   D   G   G   K   W   D  152

3091 TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA 3150
 153 Y   N   K   N   S   T   N   K   S   I   V   L   N   T   Q   S   F   T   F   E  172

3151 GAA GTA GAA TAC CTG GTT AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC GTA AAA 3210
 173 E   V   E   Y   L   V   K   G   L   R   N   K   F   Q   L   N   C   Y   V   K  192

3211 ATC AAC AAA AAC AAA CCG ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC 3270
 193 I   N   K   N   K   P   I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N  212

3271 CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC 3330
 213 L   I   K   P   Y   L   I   P   Q   M   M   Y   K   L   P   N   T   I   S   S  232

3331 GAA ACT TTC CTG AAA TAA
 233 E   T   F   L   K   *
```

2. DNA comprising the nucleotide sequence as claimed in claim 1 operatively linked to a promoter.

3. An isolated RNA sequence complementary to the nucleotide sequence of claim 1.

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
[illegible] STREET, N.W.
[illegible] WASHINGTON, D.C. 20005
202 408 4000

-continued

```
3271 CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC   3330
 213 L   I   K   P   Y   L   I   P   Q   M   M   Y   K   L   P   N   T   I   S   S    232

3331 GAA ACT TTC CTG AAA TAA
 233 E   T   F   L   K   *
```

2. DNA comprising the nucleotide sequence as claimed in claim 1 operatively linked to a promoter.

3. An isolated RNA sequence complementary to the nucleotide sequence of claim 1.

4. RNA complementary to the nucleotide sequence of claim 2.

5. A vehicle comprising a vector containing the nucleotide sequence as claimed in claim 1.

6. The vehicle as claimed in claim 5, wherein the vector is an SV-40 vector.

7. The vehicle as claimed in claim 5, wherein the vector is plasmid pSVOAL.

8. The vehicle as claimed in claim 5 having the identifying characteristics of the vector having culture collection accession number C.N.C.M. I-1014.

9. The vehicle as claimed in claim 5, wherein the vector is an expression vector.

10. A method of genetically mapping a eukaryotic genome that does not contain a natural restriction site for I-SceI, comprising the steps of:

(a) artificially inserting one or more I-SceI sites at various positions in the genome;

(b) completely cleaving said genome at the inserted I-SceI sites, with the restriction enzyme I-SceI, to produce nested chromosomal fragments;

(c) purifying said fragments of step (b) by pulsed field gel electrophoresis (PFG);

(d) transferring the fragments to a solid membrane;

(e) hybridizing the fragments bound to said membrane to a labelled probe containing DNA complementary to said fragments;

(f) detecting the hybridization banding patterns; and

(g) mapping said eukaryotic genome based on the hybridization banding patterns observed in step (f).

11. The method of claim 10, wherein said eukaryotic genome is the yeast genome.

12. The method of claim 10, wherein said eukaryotic genome is the genome of the yeast artificial chromosome vector (YAC).

13. The method of claim 10, wherein said step of artificially inserting one or more I-SceI sites comprises random insertion.

14. The method of claim 10, wherein said step of artificially inserting one or more I-SceI sites comprises homologous recombination.

15. The method of claim 11, wherein the probe of step (e) is derived from a cosmid clone, pUKG040.

16. The method of claim 11, wherein the probe of step (e) is derived from a cosmid clone, pUKG066.

17. The method of claim 10, wherein the nested chromosomal fragments of step (c) are used as hybridization probes to sort cosmid libraries.

18. The method of claim 10, wherein after step (b), the genome is partially digested with bacterial restriction enzymes of choice and then electrophoresed, as in step (c), with size calibration markers.

19. A method for in vivo site directed genetic recombination in an organism using enzyme I-SceI, comprising the steps of:

(a) introducing a synthetic gene encoding the I-SceI endonuclease into an expression vector;

(b) inserting a I-SceI restriction site next to or within a gene of interest carried on a plasmid;

(c) co-transforming the cells of said organism with said expression vector of step (a) and said plasmid of step (b), whereby said gene of interest, carried by said plasmid of step (b), is inserted into a chromosome of said organism at a specific site.

20. The method of claim 19, wherein said organism is yeast.

21. The method of claim 19, wherein said organism is bacteria.

22. The method of claim 19, wherein said organism is mouse.

23. The method of claim 19, wherein said synthetic gene of step (a) is under the control of a galactose inducible promoter.

24. The method of claim 23, wherein said expression vector is plasmid pPEX408.

25. The method of claim 23, wherein said expression vector is plasmid pPEX7.

26. A method of genetically mapping a genome that does not contain a natural restriction site for I-SceI, comprising the steps of:

(a) artificially inserting one or more I-SceI sites at various positions in the genome;

(b) completely cleaving said genome at the inserted I-SceI sites, with the restriction enzyme I-SceI, to produce nested chromosomal fragments;

(c) purifying said fragments of step (b); and

(d) mapping said eukaryotic genome by detecting said fragments.

* * * * *

Relevant Pages

ABSTRACT

An isolated DNA encoding the enzyme I-*Sce*I is provided. The DNA sequence can be incorporated in cloning and expression vectors, transformed cell lines and transgenic animals. The vectors are useful in gene mapping and site-directed insertion of genes.

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N W
WASHINGTON, DC 20005
(202 408 4000

methods for replacing a natural gene with another gene that is capable of alleviating the disease or genetic disorder.

SUMMARY OF THE INVENTION

Accordingly, this invention aids in fulfilling these needs in the art. Specifically, this invention relates to an isolated DNA encoding the enzyme I-*Sce*I. The DNA has the following nucleotide sequence:

```
                                                ATG CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG 2670
                                                M   H   M   K   N   I   K   K   N   Q   V   M   12

2671 AAC CTC GGT CCG AAC TCT AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC 2730
  13 N   L   G   P   N   S   K   L   L   K   E   Y   K   S   Q   L   I   E   L   N  32

2731 ATC GAA CAG TTC GAA GCA GGT ATC GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT 2790
  33 I   E   Q   F   E   A   G   I   G   L   I   L   G   D   A   Y   I   R   S   R  52

2791 GAT GAA GGT AAA ACC TAC TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC 2850
  53 D   E   G   K   T   Y   C   M   Q   F   E   W   K   N   K   A   Y   M   D   H  72

2851 GTA TGT CTG CTG TAC GAT CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC 2910
  73 V   C   L   L   Y   D   Q   W   V   L   S   P   P   H   K   K   E   R   V   N  92

2911 CAC CTG GGT AAC CTG GTA ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC 2970
  93 H   L   G   N   L   V   I   T   W   G   A   Q   T   F   K   H   Q   A   F   N 112

2971 AAA CTG GCT AAC CTG TTC ATC GTT AAC AAC AAA AAA ACC ATC CCG AAC AAC CTG GTT GAA 3030
 113 K   L   A   N   L   F   I   V   N   N   K   K   T   I   P   N   N   L   V   E 132

3031 AAC TAC CTG ACC CCG ATG TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT 3090
 133 N   Y   L   T   P   M   S   L   A   Y   W   F   M   D   D   G   G   K   W   D 152

3091 TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA 3150
 153 Y   N   K   N   S   T   N   K   S   I   V   L   N   T   Q   S   F   T   F   E 172

3151 GAA GTA GAA TAC CTG GTT AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC GTA AAA 3210
 173 E   V   E   Y   L   V   K   G   L   R   N   K   F   Q   L   N   C   Y   V   K 192

3211 ATC AAC AAA AAC AAA CCG ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC 3270
 193 I   N   K   N   K   P   I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N 212

3271 CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC 3330
 213 L   I   K   P   Y   L   I   P   Q   M   M   Y   K   L   P   N   T   I   S   S 232

3331 GAA ACT TTC CTG AAA TAA
 233 E   T   F   L   K   *
```

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N.W.
WASHINGTON, DC 20005
(202) 408-4000

I-*Sce*I is a double-stranded endonuclease that cleaves DNA within its recognition site. I-*Sce*I generates a 4bp staggered cut with 3'OH overhangs.

Substrate: Acts only on double-stranded DNA. Substrate DNA can be relaxed or negatively supercoiled.

Cations: Enzymatic activity requires Mg^{++} (8 mM is optimum). Mn^{++} can replace Mg^{++}, but this reduces the stringency of recognition.

Optimum conditions for activity: high pH (9 to 10), temperature 20-40°C, no monovalent cations.

Enzyme stability: I-*Sce*I is unstable at room temperature. The enzyme-substrate complex is more stable than the enzyme alone (presence of recognition sites stabilizes the enzyme.)

The enzyme I-*Sce*I has a known recognition site. (ref. 14.) The recognition site of I-*Sce*I is a non-symmetrical sequence that extends over 18 bp as determined by systematic mutational analysis. The sequence reads: (arrows indicate cuts)

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET N.W.
WASHINGTON, DC 20005
(202) 408-4000

5' TAGGGATAACAGGGTAAT 3'

3' ATCCCTATTGTCCCATTA 5'

and polymerization with the DNA polymerase. Amplified sequences can be detected by the use of a technique termed oligomer restriction (OR). See, R. K. Saiki et al., Bio/Technology 3:1008-1012 (1985).

The enzyme I-*Sce*I is one of a number of endonucleases with similar properties. Following is a listing of related enzymes and their sources.

Group I intron encoded endonucleases and related enzymes are listed below with references. Recognition sites are shown in Fig. 6.

Enzyme	Encoded by	Ref
I-*Sce*I	Sc LSU-1 intron	this work
I-*Sce*II	Sc cox1-4 intron	Sargueil et al., NAR (1990) 18, 5659-5665
I-*Sce*III	Sc cox1-3 intron	Sargueil et al., MGG (1991) 225, 340-341
I-*Sce*IV	Sc cox1-5a intron	Seraphin et al. (1992) in press
I-*Ceu*I	Ce LSU-5 intron	Marshall, Lemieux Gene (1991) 104, 241-245
I-*Cre*I	Cr LSU-1 intron	Rochaix (unpublished)
I-*Ppo*I	Pp LSU-3 intron	Muscarella et al., MCB (1990) 10, 3386-3396
I-*Tev*I	T4 td-1 intron	Chu et al., PNAS (1990) 87, 3574-3578 and Bell-Pedersen et al. NAR (1990) 18, 3763-3770.
I-*Tev*II	T4 sunY intron	Bell-Pedersen et al. NAR (1990) 18, 3763-3770.
I-*Tev*III	RB3 nrdB-1 intron	Eddy, Gold, Genes Dev. (1991) 5, 1032-1041
HO	HO yeast gene	Nickoloff et al., MCB (1990) 10, 1174-1179
Endo *Sce*I	RF3 yeast mito. gene	Kawasaki et al., JBC (1991) 266, 5342-5347

Putative new enzymes (genetic evidence but no activity as yet) are I-*Csm*I from cytochrome b intron 1 of *Chlamydomonas smithii* mitochondria (ref. 15), I-*Pan*I from cytochrome b intron 3 of *Podospora anserina* mitochondria (Jill Salvo), and

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET N W
WASHINGTON DC 20005
(202 408 4000

probably enzymes encoded by introns Nc nd1·1 and Nc cob·! from *Neurospora crassa*.

The I-endonucleases can be classified as follows:

Class I: Two dodecapeptide motifs, 4 bp staggered cut with 3′ OH overhangs, cut internal to recognition site

Subclass "I-*Sce*I"	Other subclasses
I-*Sce*I	I-*Sce*II
I-*Sce*IV	I-*Sce*III
I-*Csm*I	I-*Ceu*I (only one dodecapeptide motif)
I-*Pan*I	I-*Cre*I (only one dodecapeptide motif)
	HO
	TFP1-408 (HO homolog)
	Endo *Sce*I

Class II: GIY-(N_{10-11}) YIG motif, 2 bp staggered cut with 3′ OH overhangs, cut external to recognition site:

I-*Tev*I

Class III: no typical structural motifs, 4 bp staggered cut with 3′ OH overhangs, cut internal to recognition site:

I-*Ppo*I

Class IV: no typical structural motifs, 2 bp staggered cut with 3′ OH overhangs, cut external to recognition site:

I-*Tev*II

Class V: no typical structural motifs, 2 bp staggered cut with 5′ OH overhangs:

I-*Tev*III.

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N. W.
WASHINGTON, DC 20005
(202) 408-4000

York (1982). The nucleic acid will generally be obtained from a bacteria, yeast, virus, or a higher organism, such as a plant or animal. The nucleic acid can be a fraction of a more complex mixture, such as a portion of a gene contained in whole human DNA or a portion of a nucleic acid sequence of a particular microorganism. The nucleic acid can be a fraction of a larger molecule or the nucleic acid can constitute an entire gene or assembly of genes. The DNA can be in a single-stranded or double-stranded form. If the fragment is in single-stranded form, it can be converted to double-stranded form using DNA polymerase according to conventional techniques.

The DNA sequence of the invention can be linked to a structural gene. As used herein, the term "structural gene" refers to a DNA sequence that encodes through its template or messenger mRNA a sequence of amino acids characteristic of a specific protein or polypeptide. The nucleotide sequence of the invention can function with an expression control sequence, that is, a DNA sequence that controls and regulates expression of the gene when operatively linked to the gene.

4. Vectors Containing the Nucleotide Sequence of the Invention

This invention also relates to cloning and expression vectors containing the DNA sequence of the invention coding for the enzyme I-*Sce*I.

More particularly, the DNA sequence encoding the enzyme can be ligated to a vehicle for cloning the sequence. The

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
300 I STREET, N.W.
WASHINGTON, D.C. 20005
(202) 408-4000

Examples of yeast expression vectors are:

pPEX7, which is a yeast expression vector derived from pRP51-Bam O (a LEU2d derivative of pLG-SD5) by insertion of the synthetic gene under the control of the galactose promoter. Expression is induced by galactose.

pPEX408, which is a yeast expression vector derived from pLG-SD5 by insertion of the synthetic gene under the control of the galactose promoter. Expression is induced by galactose.

Several yeast expression vectors are depicted in Fig. 7.

Typical mammalian expression vectors are:

pRSV I-*Sce*I, which is a pRSV derivative in which the synthetic gene (*Bam*HI - *Pst*I fragment from pSCM525) is under the control of the LTR promoter of Rous Sarcoma Virus. This expression vector is depicted in Fig. 8.

Vectors for expression in Chinese Hamster Ovary (CHO) cells can also be employed.

5. Cells Transformed with Vectors of the Invention

The vectors of the invention can be inserted into host organisms using conventional techniques. For example, the vectors can be inserted by transformation, transfection, electroporation, microinjection, or by means of liposomes (lipofection).

Cloning can be carried out in prokaryotic or eukaryotic cells. The host for replicating the cloning vehicle will of course be one that is compatible with the vehicle and in

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
300 I STREET, N W
WASHINGTON DC 20005
(202 408 4000

hybridization between the probe pUKG066 and chromosome III, probably due to some repetitive DNA sequences.

All chromosome fragments, taken together, now define physical intervals as indicated in Fig. 15d. The I-*Sce*I map obtained has an 80 kb average resolution.

Example 2: Application of the Nested Chromosomal Fragmentation Strategy to the Mapping of Yeast Artificial Chromosome (YAC) Clones

This strategy can be applied to YAC mapping with two possibilities.

-1- insertion of the I-*Sce*I site within the gene of interest using homologous recombination in yeast. This permits mapping of that gene in the YAC insert by I-*Sce*I digestion *in vitro*. This has been done and works.

-2- random integration of I-*Sce*I sites along the YAC insert by homologous recombination in yeast using highly repetitive sequences (e.g., B2 in mouse or Alu in human). Transgenic strains are then used as described in ref. P1 to sort libraries or map genes.

The procedure has now been extended to YAC containing 450 kb of Mouse DNA. To this end, a repeated sequence of mouse DNA (called B2) has been inserted in a plasmid containing the I-*Sce*I site and a selectable yeast marker (LYS2). Transformation of the yeast cells containing the recombinant YAC with the plasmid linearized within the B2 sequence resulted in the integration of the I-*Sce*I site at five different locations distributed along the mouse DNA insert. Cleavage at the inserted I-*Sce*I sites using the

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N. W.
WASHINGTON, DC 20005
(202) 408-4000

7. B. Dujon, G. Cottarel, L. Colleaux, M. Betermier, A. Jacquier, L. D'Auriol, F. Galibert, Mechanism of integration of an intron within a mitochondrial gene: a double strand break and the transposase function of an intron encoded protein as revealed by *in vivo* and *in vitro* assays. In Achievements and perspectives of Mitochondrial Research". Vol. II, Biogenesis, E. Quagliariello et al. Eds. Elsevier, Amsterdam (1985) pages 215-225.

8. L. Colleaux, L. D'Auriol, M. Betermier, G. Cottarel, A. Jacquier, F. Galibert, and B. Dujon, A universal code equivalent of a yeast mitochondrial intron reading frame is expressed into Escherichia coli as a specific double strand endonuclease. Cell (1986) 44, 521-533.

9. B. Dujon, L. Colleaux, A. Jacquier, F. Michel and C. Monteilhet, Mitochondrial introns as mobile genetic elements: the role of intron-encoded proteins. In "Extrachromosomal elements in lower eucaryotes", Reed B et al. Eds. (1986) Plenum Pub. Corp. 5-27.

10. F. Michel and B. Dujon, Genetic Exchanges between Bacteriophage T4 and Filamentous Fungi? Cell (1986) 46, 323.

11. L. Colleaux, L. D'Auriol, F. Galibert and B. Dujon, Recognition and cleavage site of the intron encoded omega transposase. PNAS (1988), 85, 6022-6026.

12. B. Dujon, Group I introns as mobile genetic elements, facts and mechanistic speculations: A Review. Gene (1989), 82, 91-114.

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N.W.
WASHINGTON, D.C. 20005
(202) 408-4000

under the control of an inducible promoter can be constructed.

2. Effects of site specific double strand breaks in yeast (refs. 18 and P4)

Effects on plasmid-borne I-*Sce*I sites:

Intramolecular effects are described in detail in Ref. 18. Intermolecular (plasmid to chromosome) recombination can be predicted.

stray copy mark (should not have been printed as a period).

Effects on chromosome integrated I-*Sce*I sites

In a haploid cell, a single break within a chromosome at an artificial I-*Sce*I site results in cell division arrest followed by death (only a few % of survival). Presence of an intact sequence homologous to the cut site results in repair and 100% cell survival. In a diploid cell, a single break within a chromosome at an artificial I-*Sce*I site results in repair using the chromosome homolog and 100% cell survival. In both cases, repair of the induced double strand break results in loss of heterozygosity with deletion of the non homologous sequences flanking the cut and insertion of the non homologous sequences from the donor DNA molecule.

3. Application for *in vivo* recombination YACs in Yeast

Construction of a YAC vector with the I-*Sce*I restriction site next to the cloning site should permit one to induce homologous recombination with another YAC if inserts are

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N.W.
WASHINGTON, DC 20005
(202) 408-4000

13. B. Dujon, M. Belfort, R.A. Butow, C.Jacq, C. Lemieux, P.S. Perlman, V.M. Vogt, Mobile introns: definition of terms and recommended nomenclature. Gene (1989), 82, 115-118.

14. C. Monteilhet, A. Perrin, A. Thierry, L. Colleaux, B. Dujon, Purification and Characterization of the *in vitro* activity of I-*Sce*I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acid Research (1990), 18, 1407-1413.

15. L. Colleaux, M-R. Michel-Wolwertz, R.F. Matagne, B. Dujon - The apocytochrome b gene of Chlamydomonas smithii contains a mobile intron related to both Saccharomyces and Neurospora introns. Mol. Gen. Genet. (1990) 223, 288-296.

16. B. Dujon Des introns autonomes et mobiles. Annales de I'Institut Pasteur/ Actualites (1990) 1. 181-194.

17. A. Thierry, A. Perrin, J. Boyer, C. Fairhead, B. Dujon, B. Frey, G. Schmitz. Cleavage of yeast and bacteriophage 17 genomes at a single site using the rare cutter endonuclease I-*Sce*. I Nuc. Ac. Res. (1991) 19, 189-190.

18. A. Plessis, A. Perrin, J.E. Haber, B. Dujon, Site specific recombination determined by I-*Sce*I, a mitochondrial intron-encoded endonuclease expressed in the yeast nucleus. GENETICS (1992) 130, 451-460.

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N.W.
WASHINGTON, DC 20005
(202) 408-4000

A11. B. Dujon, A. Jacquier, L. Colleaux, C. Monteilhet, A. Perrin, "Les Introns autoepissables et leurs proteins" Colloque "Biologie Moleculaire de la levure: expression genetique chez Saccharomyces "organise par la Societe francaise de Microbiologie 18 January 1988 Institut Pasteur, Paris.

A12. L. Colleaux, L. D'Auriol, C. Monteilhet, F. Galibert and B. Dujon, Characterization of the biochemical activity of an intron encoded transposase. 14th International Conference on Yeast Genetics and Molecular Biology. Espoo, Finland, 7-13 August 1988.

A13. B. Dujon, A goup I intron as a mobile genetic element, Albany Conference sur "RNA: catalysis, splicing, evolution", Albany, New York, 22-25 September 1988.

A14. B. Dujon, L. Colleaux, C. Monteilhet, A. Perrin, L. D'Auriol, F. Galibert, Group I introns as mobile genetic elements: the role of intron encoded proteins and the nature of the target site. 14th Annual EMBO Symposium "Organelle genomes and the nucleus" Heidelberg, 26-29 September 1988.

A15. L. Colleaux, R. Matagne, B. Dujon, A new mobile mitochondrial intron provides evidence for genetic exchange between Neurospora and Chlamydomonas species. Cold Spring Harbor, May 1989.

A16. L. Colleaux, M.R. Michel-Wolwertz, R.F. Matagne, B. Dujon, The apoxytochrome b gene of Chlamydomonas smithii contains a mobile intron related to both Saccharomyces and Neurospora introns. Fourth International Conference

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N.W.
WASHINGTON, DC 20005
(202) 408-4000

on Cell and Molecular Biology of Chlamydomonas. Madison, WI, April 1990.

A17. B. Dujon, L. Colleaux, E. Luzi, C. Monteilhet, A. Perrin, A. Plessis, I. Stroke, A. Thierry, Mobile Introns, EMBO Workshop on "Molecular Mechanisms of transposition and its control, Roscoff (France) June 1990.

A18. A. Perrin, C. Monteilhet, A. Thierry, E. Luzi, I. Stroke, L. Colleaux, B. Dujon. I-*Sce*I, a novel double strand site specific endonuclease, encoded by a mobile group I intron in Yeast. Workshop on "RecA and Related Proteins" Sacly, France 17-21 September 1990.

A19. A. Plessis, A. Perrin, B. Dujon, Site specific recombination induced by double strand endonucleases, HO and I-*Sce*I in yeast. Workshop on "RecA and Related Proteins" Saclay, France 17-21 September 1990.

A20. B. Dujon, The genetic propagation of introns 20th FEBS Meeting, Budapest, Hungary, August 1990.

A21. E. Luzi, B. Dujon, Analysis of the intron encoded site specific endonuclease I-*Sce*I by mutagenesis, Third European Congress on Cell Biology, Florence, Italy, September 1990.

A22. B. Dujon, Self splicing introns as contagious genetic elements. Journees Franco-Beiges de Pont a Mousson. October 1990.

A23. B. Frey, H. Dubler, G. Schmitz, A. Thierry, A. Perrin, J. Boyer, C. Fairhead, B. Dujon, Specific cleavage of the yeast genome at a single site using the rare cutter

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET N.W.
WASHINGTON, D.C. 20005
(202) 408-4000

endonuclease I-*Sce*I Human Genome, Frankfurt, Germany, November 1990.

A24. B. Dujon, A. Perrin, I. Stroke, E. Luzi, L. Colleaux, A. Plessis, A. Thierry, The genetic mobility of group I introns at the DNA level. Keystone Symposia Meeting on "Molecular Evolution of Introns and Other RNA elements", Taos, New Mexico, 2-8 February 1991.

A25. B. Dujon, J. Boyer, C. Fairhead, A. Perrin, A Thierry, Cartographie chez la levure. Reunion "Strategies d'etablissement des cartes geniques" Toulouse 30-31 Mai 1991.

A26. B. Dujon, A. Thierry, Nested chromosomal fragmentation using the meganuclease I-*Sce*I: a new method for the rapid mapping of the yeast genome. Elounda, Crete 15-17 Mail 1991.

A27. A. Thierry, L. Gaillon, F. Galibert, B. Dujon. The chromosome XI library: what has been accomplished, what is left. Brugge meeting 22-24 September 1991.

A28. B. Dujon, A. Thierry, Nested chromosomal fragmentation using the meganuclease I-*Sce*I: a new method for the rapid physical mapping of the eukaryotic genomes. Cold Spring Harbor 6-10 May 1992.

A29. A. Thierry, L. Gaillon, F. Galibert, B. Dujon. Yeast chromosome XI: construction of a cosmid contig. a high resolution map and sequencing progress. Cold Spring Harbor 6-10 May 1992.

TTTCACACAG GAAACAGCTA TGACCATGAT TACGAATTCT CATGTTTGAC AGCTTATCAT 600

CGATAAGCTT TAATGCGGTA GTTTATCACA GTTAAATTGC TAACGCAGTC AGGCACCGTG 660

TATGAAATCT AACAATGCGC TCATCGTCAT CCTCGGCACC GTCACCCTGG ATGCTGTAGG 720

CATAGGCTTG GTTATGCCGG TACTGCCGGG CCTCTTGCGG GATATCCGCC TGATGCGTGA 780

ACGTGACGGA CGTAACCACC GCGACATGTG TGTGCTGTTC CGCTGGGCAT GCCAGGACAA 840

CTTCTGGTCC GGTAACGTGC TGAGCCCGGC CAAGCTTACT CCCCATCCCC CTGTTGACAA 900

TTAATCATCG GCTCGTATAA TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA 960

CAGGATCCAT GCATATGAAA AACATCAAAA AAAACCAGGT AATGAACCTG GGTCCGAACT 1020

CTAAACTGCT GAAAGAATAC AAATCCCAGC TGATCGAACT GAACATCGAA CAGTTCGAAG 1080

CAGGTATCGG TCTGATCCTG GGTGATGCTT ACATCCGTTC TCGTGATGAA GGTAAAACCT 1140

ACTGTATGCA GTTCGAGTGG AAAAACAAAG CATACATGGA CCACGTATGT CTGCTGTACG 1200

ATCAGTGGGT ACTGTCCCCG CCGCACAAAA AAGAACGTGT TAACCACCTG GGTAACCTGG 1260

TAATCACCTG GGGCGCCCAG ACTTTCAAAC ACCAAGCTTT CAACAAACTG GCTAACCTGT 1320

TCATCGTTAA CAACAAAAAA ACCATCCCGA ACAACCTGGT TGAAAACTAC CTGACCCCGA 1380

TGTCTCTGGC ATACTGGTTC ATGGATGATG GTGGTAAATG GGATTACAAC AAAAACTCTA 1440

CCAACAAATC GATCGTACTG AACACCCAGT CTTTCACTTT CGAAGAAGTA GAATACCTGG 1500

TTAAGGGTCT GCGTAACAAA TTCCAACTGA ACTGTTACGT AAAAATCAAC AAAAACAAAC 1560

CGATCATCTA CATCGATTCT ATGTCTTACC TGATCTTCTA CAACCTGATC AAACCGTACC 1620

TCATCCCCCA GATGATGTAC AAACTGCCGA ACACTATCTC CTCCGAAACT TTCCTGAAAT 1680

AATAAGTCGA CCTGCAGCCC AAGCTTGGCA CTGGCCGTCG TTTTACAACG TCGTGACT 1738

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

What is claimed is:

1. An isolated DNA encoding the enzyme I-*Sce*I, wherein the DNA has the nucleotide squence:

```
                                                ATG CAT ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG 2670
                                                M   H   M   K   N   I   K   K   N   Q   V   M   12

2671 AAC CTC GGT CCG AAC TCT AAA CTG CTG AAA GAA TAC AAA TCC CAG CTG ATC GAA CTG AAC 2730
  13 N   L   G   P   N   S   K   L   L   K   E   Y   K   S   Q   L   I   E   L   N   32

2731 ATC GAA CAG TTC GAA GCA GGT ATC GGT CTG ATC CTG GGT GAT GCT TAC ATC CGT TCT CGT 2790
  33 I   E   Q   F   E   A   G   I   G   L   I   L   G   D   A   Y   I   R   S   R   52

2791 GAT GAA GGT AAA ACC TAC TGT ATG CAG TTC GAG TGG AAA AAC AAA GCA TAC ATG GAC CAC 2850
  53 D   E   G   K   T   Y   C   M   Q   F   E   W   K   N   K   A   Y   M   D   H   72

2851 GTA TGT CTG CTG TAC GAT CAG TGG GTA CTG TCC CCG CCG CAC AAA AAA GAA CGT GTT AAC 2910
  73 V   C   L   L   Y   D   Q   W   V   L   S   P   P   H   K   K   E   R   V   N   92

2911 CAC CTG GGT AAC CTG GTA ATC ACC TGG GGC GCC CAG ACT TTC AAA CAC CAA GCT TTC AAC 2970
  93 H   L   G   N   L   V   I   T   W   G   A   Q   T   F   K   H   Q   A   F   N   112

2971 AAA CTG GCT AAC CTG TTC ATC GTT AAC AAC AAA AAA ACC ATC CCG AAC AAC CTG GTT GAA 3030
 113 K   L   A   N   L   F   I   V   N   N   K   K   T   I   P   N   N   L   V   E   132

3031 AAC TAC CTG ACC CCG ATG TCT CTG GCA TAC TGG TTC ATG GAT GAT GGT GGT AAA TGG GAT 3090
 133 N   Y   L   T   P   M   S   L   A   Y   W   F   M   D   D   G   G   K   W   D   152

3091 TAC AAC AAA AAC TCT ACC AAC AAA TCG ATC GTA CTG AAC ACC CAG TCT TTC ACT TTC GAA 3150
 153 Y   N   K   N   S   T   N   K   S   I   V   L   N   T   Q   S   F   T   F   E   172

3151 GAA GTA GAA TAC CTG GTT AAG GGT CTG CGT AAC AAA TTC CAA CTG AAC TGT TAC GTA AAA 3210
 173 E   V   E   Y   L   V   K   G   L   R   N   K   F   Q   L   N   C   Y   V   K   192

3211 ATC AAC AAA AAC AAA CCG ATC ATC TAC ATC GAT TCT ATG TCT TAC CTG ATC TTC TAC AAC 3270
 193 I   N   K   N   K   P   I   I   Y   I   D   S   M   S   Y   L   I   F   Y   N   212

3271 CTG ATC AAA CCG TAC CTG ATC CCG CAG ATG ATG TAC AAA CTG CCG AAC ACT ATC TCC TCC 3330
 213 L   I   K   P   Y   L   I   P   Q   M   M   Y   K   L   P   N   T   I   S   S   232

3331 GAA ACT TTC CTG AAA TAA
 233 E   T   F   L   K   *
```

2. DNA comprising the nucleotide sequence as claimed in claim 1 operatively linked to a promoter.

3. An isolated RNA sequence complementary to the nucleotide sequence of claim 1.

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
300 I STREET, N.W.
WASHINGTON, D.C. 20005
202 408 4000

(f) detecting the hybridization banding patterns; and

(g) mapping said eukaryotic genome based on the hybridization banding patterns observed in step (f).

11. The method of claim 10, wherein said eukaryotic genome is the yeast genome.

12. The method of claim 10, wherein said eukaryotic genome is the genome of the yeast artificial chromosome vector (YAC).

13. The method of claim 10, wherein said step of artificially inserting one or more I-*Sce*I sites comprises random insertion.

14. The method of claim 10, wherein said step of artificially inserting one or more I-*Sce*I sites comprises homologous recombination.

15. The method of claim 11, wherein the probe of step (e) is derived from a cosmid clone, pUKG040.

16. The method of claim 11, wherein the probe of step (e) is derived from a cosmid clone, pUKG066.

17. The method of claim 10, wherein the nested chromosomal fragments of step (c) are used as hybridization probes to sort cosmid libraries.

18. The method of claim 10, wherein after step (b), the genome is partially digested with bacterial restriction enzymes of choice and then electrophoresed, as in step (c), with size calibration markers.

19. A method for *in vivo* site directed genetic recombination in an organism using enzyme I-*Sce*I, comprising the steps of: